CHAPTER 2. Cramer-Rao lower bound

- Given an estimation problem, what is the variance of the best possible estimator?
- This quantity is given by the *Cramer-Rao lower bound* (CRLB), which we will study in this section.
- As a side product, the CRLB theorem gives also a method for finding the best estimator.
- However, this is not very practical. There exists easier methods, which will be studied in later chapters.

Discovering the CRLB idea

- Consider again the problem of estimating the DC level A in the model
$$x[n] = A + w[n],$$
where $w[n] \sim \mathcal{N}(0, \sigma^2)$.
- For simplicity, suppose that we're using only a single observation to do this: $x[0]$.
- Now the pdf of $x[0]$ is

$$p(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[0] - A)^2\right]$$

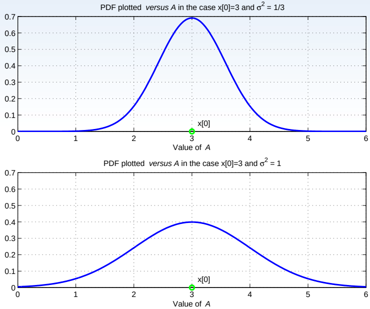
Discovering the CRLB idea (cont.)

- Once we have observed $x[0]$, say for example $x[0] = 3$, some values of A are more likely than others.
- Actually, the pdf of A has the same form as the PDF of $x[0]$:

$$\text{pdf of } A = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(3 - A)^2\right]$$

This pdf is plotted below for two values of σ^2.

Discovering the CRLB idea (cont.)



- Observation: in the upper case we will probably get more accurate results.

Discovering the CRLB idea

- If the pdf is viewed as a function of the unknown parameter (with **x** fixed), it is called the *likelihood function*.
- Often the likelihood function has an exponential form. Then it's usual to take the natural logarithm to get rid of the exponential. Note that the maximum of the new *log-likelihood* function does not change.
- The "sharpness" of the likelihood function determines how accurately we can estimate the parameter. For example the pdf on top is the easier case. If a single sample is observed as $x[0] = A + w[0]$, then we can expect a better estimate if σ^2 is small.

Discovering the CRLB idea (cont.)

- How to quantify the sharpness?
  - We could try to guess a good estimator, and find it's variance
  - But this gives only information of our estimator, not of the whole estimation problem (and all possible estimators).
- Is there any measure that would be common for all possible estimators for a specific estimation problem? Specifically, we'd like to find the smallest possible variance.
- The second derivative of the likelihood function (or log-likelihood) is one alternative for measuring the sharpness of a function.

Discovering the CRLB idea (cont.)

- Let's see what it is in our simple case. Note, that in this case the minimum variance of all estimators is σ^2, since we're using only one sample (we can't prove it, but it seems obvious).
- From the likelihood function

$$p(x[0];A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[0]-A)^2\right]$$

we get the log-likelihood function:

$$\ln p(x[0];A) = -\ln\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(x[0]-A)^2$$

Discovering the CRLB idea (cont.)

- The derivative with respect to A is:

$$\frac{\partial \ln p(x[0]; A)}{\partial A} = \frac{1}{\sigma^2}(x[0] - A)$$

- And the second derivative:

$$\frac{\partial^2 \ln p(x[0]; A)}{\partial A^2} = -\frac{1}{\sigma^2}$$

Discovering the CRLB idea (cont.)

- Since σ^2 is the smallest possible variance, in this case we have an alternative way of finding the minimum variance of all estimators:

$$\text{minimum variance} = \sigma^2 = \frac{1}{-\frac{\partial^2 \ln p(x[0];A)}{\partial A^2}}$$

- Does our hypothesis hold in general?

Discovering the CRLB idea (cont.)

- The answer is yes, with the following notes.
  - Note 1: if the function depends on the data $\mathbf{x}$, take the expectation over all $\mathbf{x}$.
  - Note 2: if the function depends on the parameter θ, evaluate the derivative at the true value of θ.
- Thus, we have the following rule:

$$\text{minimum variance of any unbiased estimator} = \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]}$$

- The following two slides state the above in mathematical terms. Additionally, there is a rule for finding an estimator reaching the CRLB.

Cramer Rao Lower Bound - theorem

Theorem: CRLB - scalar parameter
It is assumed that the pdf $p(\mathbf{x};\theta)$ *satisfies the "regularity" condition*

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0, \quad \forall\theta \tag{1}$$

where the expectation is taken with respect to $p(\mathbf{x};\theta)$. *Then, the variance of any unbiased estimator* $\hat{\theta}$ *must satisfy*

$$var(\hat{\theta}) \geqslant \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]} \tag{2}$$

Cramer Rao Lower Bound - theorem (cont.)

where the derivative is evaluated at the true value of θ *and the expectation is taken with respect to* $p(\mathbf{x};\theta)$. *Furthermore, an unbiased estimator may be found that attains the bound for all* θ *if and only if*

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} = I(\theta)(g(x) - \theta) \tag{3}$$

for some functions g *and* I. *That estimator, which is the MVU estimator is* $\hat{\theta} = g(x)$ *and the minimum variance is* $1/I(\theta)$.

Proof. Three pages of integrals; see appendix 3A of Kay's book.

CRLB Example 1: estimation of DC level in WGN

- *Example* 1. DC level in white Gaussian noise with N data points:

$$x[n] = A + w[n], \quad n = 0, 1, \dots, N-1$$

What's the minimum variance of any unbiased estimator using N *samples?*

CRLB Example 1: estimation of DC level in WGN (cont.)

- Now the pdf (and the likelihood function) is the product of individual densities:

$$
\begin{aligned}
p(\mathbf{x}; A) &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[n] - A)^2\right] \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A)^2\right].
\end{aligned}
$$

CRLB Example 1: estimation of DC level in WGN (cont.)

- The log-likelihood function is now

$$\ln p(\mathbf{x}; A) = -\ln[(2\pi\sigma^2)^{\frac{N}{2}}] - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A)^2.$$

- The first derivative is

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x}, A)}{\partial A} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n] - A)\\ &= \frac{N}{\sigma^2}(\bar{\mathbf{x}} - A),\end{aligned}$$

where $\bar{\mathbf{x}}$ denotes the sample mean.

CRLB Example 1: estimation of DC level in WGN (cont.)

- The second derivative has a simple form:

$$\frac{\partial^2 \ln p(\mathbf{x}, A)}{\partial A^2} = -\frac{N}{\sigma^2}$$

- Therefore, the minimum variance of any unbiased estimator is

$$\mathrm{var}(\hat{A}) \geqslant \frac{\sigma^2}{N}$$

- In lecture 1 we saw that this variance can be achieved using the sample mean estimator. Therefore, the sample mean is a MVUE for this problem—no better estimators exist.

CRLB Example 1: estimation of DC level in WGN (cont.)

- Note that we could have arrived at the optimal estimator also using the CRLB theorem, which says that an unbiased estimator attaining the CRLB exists if and only if $\frac{\partial \ln p(\mathbf{x}, A)}{\partial A}$ can be factored as

$$\frac{\partial \ln p(\mathbf{x}, A)}{\partial A} = I(A)(g(\mathbf{x}) - A).$$

CRLB Example 1: estimation of DC level in WGN (cont.)

- Earlier we saw that

$$\frac{\partial \ln p(\mathbf{x}, A)}{\partial A} = \frac{N}{\sigma^2}(\bar{\mathbf{x}} - A),$$

which means that this factorization exists with

$$I(A) = \frac{N}{\sigma^2} \text{ and } g(\mathbf{x}) = \bar{\mathbf{x}}.$$

Furthermore, $g(\mathbf{x})$ is the MVUE and $1/I(A)$ is its variance.

CRLB Example 2: estimation of the phase of a sinusoid

- *Example* 2. Phase estimation: estimate the phase ϕ of a sinusoid embedded in WGN. Suppose that the amplitude A and frequency f_0 are known.
- We measure N data points assuming the model:

$$x[n] = A\cos(2\pi f_0 n + \phi) + w[n], \quad n = 0, 1, \ldots, N-1$$

How accurately is it possible to estimate ϕ? Can we find this estimator?

CRLB Example 2: estimation of the phase of a sinusoid (cont.)

- Let's start with the pdf:

$$p(\mathbf{x};\phi) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}[x[n] - A\cos(2\pi f_0 n + \phi)]^2\right\}$$

CRLB Example 2: estimation of the phase of a sinusoid (cont.)

- The derivative of its logarithm is

$$
\begin{aligned}
&\frac{\partial \ln p(\mathbf{x};\phi)}{\partial \phi} \\
&= -\frac{1}{\sigma^2}\sum_{n=0}^{N-1}[x[n] - A\cos(2\pi f_0 n + \phi)] \cdot A\sin(2\pi f_0 n + \phi) \\
&= -\frac{A}{\sigma^2}\sum_{n=0}^{N-1}\left[x[n]\sin(2\pi f_0 n + \phi) - A\underbrace{\cos(2\pi f_0 n + \phi)\sin(2\pi f_0 n + \phi)}_{\frac{1}{2}\sin(2\cdot(2\pi f_0 n + \phi))}\right] \\
&= -\frac{A}{\sigma^2}\sum_{n=0}^{N-1}\left[x[n]\sin(2\pi f_0 n + \phi) - \frac{A}{2}\sin(4\pi f_0 n + 2\phi)\right]
\end{aligned}
$$

CRLB Example 2: estimation of the phase of a sinusoid (cont.)

- And the second derivative:

$$\frac{\partial^2 \ln p(\mathbf{x};\phi)}{\partial \phi^2} = -\frac{A}{\sigma^2}\sum_{n=0}^{N-1}[x[n]\cos(2\pi f_0 n+\phi)-A\cos(4\pi f_0 n+2\phi)]$$

CRLB Example 2: estimation of the phase of a sinusoid (cont.)

- This time the second derivative depends on the data $\mathbf{x}$. Therefore, we have to take the expectation:

$$
\begin{aligned}
& E\left[\frac{\partial^2 \ln p(\mathbf{x};\phi)}{\partial \phi^2}\right] \\
= \; & E\left[-\frac{A}{\sigma^2}\sum_{n=0}^{N-1}[x[n]\cos(2\pi f_0 n + \phi) - A\cos(4\pi f_0 n + 2\phi)]\right] \\
= \; & -\frac{A}{\sigma^2}\sum_{n=0}^{N-1}[E[x[n]]\cos(2\pi f_0 n + \phi) - A\cos(4\pi f_0 n + 2\phi)]
\end{aligned}
$$

CRLB Example 2: estimation of the phase of a sinusoid (cont.)

- The expectation simplifies to

$$E[x[n]] = A\cos(2\pi f_0 n + \phi) + E[w[n]] = A\cos(2\pi f_0 n + \phi),$$

and thus we have the following form

$$\begin{aligned} E\left[\frac{\partial^2 \ln p(\mathbf{x};\phi)}{\partial \phi^2}\right] &= -\frac{A}{\sigma^2}\sum_{n=0}^{N-1}\left[A\cos^2(2\pi f_0 n + \phi) - A\cos(4\pi f_0 n + 2\phi)\right] \\ &= -\frac{A^2}{\sigma^2}\sum_{n=0}^{N-1}\left[\cos^2(2\pi f_0 n + \phi) - \cos(4\pi f_0 n + 2\phi)\right]. \end{aligned}$$

CRLB Example 2: estimation of the phase of a sinusoid (cont.)

- We can evaluate this approximately by noting that the squared cosine has the form

$$\cos^2(x) = \frac{1}{2} + \frac{1}{2}\cos(2x).$$

CRLB Example 2: estimation of the phase of a sinusoid (cont.)

- Therefore,

$$
\begin{aligned}
& E\left[\frac{\partial^2 \ln p(\mathbf{x};\phi)}{\partial\phi^2}\right] \\
= \quad & -\frac{A^2}{\sigma^2}\sum_{n=0}^{N-1}\left[\frac{1}{2}+\frac{1}{2}\cos(4\pi f_0 n+2\phi)-\cos(4\pi f_0 n+2\phi)\right] \\
= \quad & -\frac{A^2}{\sigma^2}\sum_{n=0}^{N-1}\left[\frac{1}{2}-\frac{1}{2}\cos(4\pi f_0 n+2\phi)\right] \\
= \quad & -\frac{NA^2}{2\sigma^2}+\frac{A^2}{2\sigma^2}\sum_{n=0}^{N-1}\cos(4\pi f_0 n+2\phi)
\end{aligned}
$$

CRLB Example 2: estimation of the phase of a sinusoid (cont.)

- As N grows, the effect of the sum of cosines becomes negligible compared to the first term (unless $f_0 = 0$ or $f_0 = \frac{1}{2}$, which are pathological cases anyway).
- Therefore, we can approximate roughly:

$$\mathrm{var}(\hat{\phi}) \underset{\text{approx.}}{\geqslant} -\frac{1}{-\frac{NA^2}{2\sigma^2}} = \frac{2\sigma^2}{NA^2}$$

for any unbiased estimator $\hat{\phi}$

CRLB Example 2: estimation of the phase of a sinusoid (cont.)

- Note that in this case the factorization of CRLB theorem can not be done. This means that an unbiased estimator achieving the bound does not exist. However, a MVUE may still exist.
- Note also that we could have evaluated the exact boundary using computer for a fixed f_0. Now we approximated to get a nice looking formula.

CRLB summary

- Cramer Rao inequality provides lower bound for the estimation error variance.
- Minimum attainable variance is often larger than CRLB.
- We need to know the pdf to evaluate CRLB. Often we don't know this information and cannot evaluate this bound. If the data is multivariate Gaussian or i.i.d. with known distribution, we can evaluate it.[1],
- If the estimator reaches the CRLB, it is called *efficient*.
- MVUE may or may not be efficient. If it is not, we have to use other tools than CRLB to find it (see Chapter 5).
- It's not guaranteed that MVUE exists or is realizable.

[1] i.i.d. is a widely used abbreviation for *independent and identically distributed*. This means that such random variables all have the same probability distribution and they are mutually independent.

Fisher information

- The CRLB states that

$$\mathrm{var}(\hat{\theta}) \geqslant \frac{1}{-\mathrm{E}\left[\frac{\partial^2 \ln p(x[0];\theta)}{\partial \theta^2}\right]}$$

- The denominator is so significant, that it has its own name: *Fisher information*, and is denoted by

$$I(\theta) = -\mathrm{E}\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]$$

Fisher information (cont.)

- In the proof of CRLB, $I(\theta)$ has also another form:

$$I(\theta) = E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right)^2\right]$$

- The Fisher information is
  1. nonnegative
  2. additive for independent observations, i.e., if the samples are i.i.d.,

$$I(\theta) = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right] = -\sum_{n=0}^{N-1} E\left[\frac{\partial^2 \ln p(x[n];\theta)}{\partial \theta^2}\right]$$

CRLB for a general WGN case

- Since we often assume that the signal is embedded in white Gaussian noise, it is worthwhile to derive a general formula of CRLB in this case.
- Suppose, that a deterministic signal depending on an unknown parameter θ in WGN is observed:

$$x[n] = s[n] + w[n], \qquad n = 0, 1, \ldots, N-1,$$

where only $s[n]$ depends on θ.

CRLB for a general WGN case (cont.)

- It can be shown, that the variance of any unbiased estimator $\hat{\theta}$

$$\operatorname{var}(\hat{\theta}) \geqslant \frac{\sigma^2}{\sum_{n=0}^{N-1} \left(\frac{\partial s[n]}{\partial \theta} \right)^2}$$

- Note that if $s[n]$ is constant, we have the familiar DC level in WGN case, and $\operatorname{var}(\hat{\theta}) \geqslant \sigma^2/N$, as expected.

CRLB for a general WGN case—example

- Suppose we observe $x[n]$ assuming the model

$$x[n] = A\cos(2\pi f_0 n + \phi) + w[n],$$

where A and ϕ are know, and f_0 is the unkown parameter to be estimated. Moreover, $w[n] \sim \mathcal{N}(0, \sigma^2)$.

- Using the general rule for CRLB in WGN, we only need to solve

$$\sum_{n=0}^{N-1} \left(\frac{\partial s[n]}{\partial f_0} \right)^2$$

CRLB for a general WGN case—example (cont.)

- Now, the derivative equals

$$\frac{\partial s[n]}{\partial f_0} = \frac{\partial}{\partial f_0}(A\cos(2\pi f_0 n + \phi)) = -2\pi A n \sin(2\pi f_0 n + \phi).$$

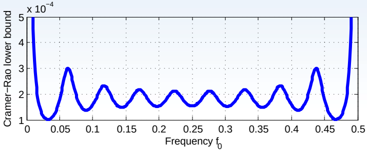
- Thus, the variance of any unbiased estimator $\hat{f}_0$

$$\mathrm{var}(\hat{f}_0) \geqslant \frac{\sigma^2}{(2\pi A)^2 \sum_{n=0}^{N-1} n^2 \sin^2(2\pi f_0 n + \phi)}$$

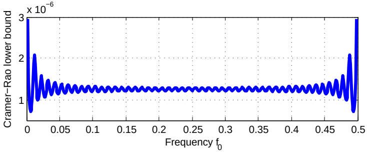
- The plot of this bound for different frequencies is shown below.

CRLB for a general WGN case—example (cont.)

- Here $N = 10$, $\phi = 0$, $A^2 = \sigma^2 = 1$.



- Below is the case with $N = 50$.



- There exists a number of "easy" frequencies.

CRLB - Transformation of parameters

- Suppose we want to estimate $\alpha = g(\theta)$ through estimating θ first. Is it enough just to apply the transformation to the estimation result?
- As an example, instead of estimating the DC level A of WGN:

$$x[n] = A + w[n],$$

  we might be interested in estimating the power A^2. What is the CRLB for any estimator of A^2?

CRLB - Transformation of parameters (cont.)

- It can be shown (Kay: Appendix 3A) that the CRLB is

$$\operatorname{var}(\hat{\alpha}) \geqslant \frac{\left(\frac{\partial g}{\partial \theta}\right)^2}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]}.$$

- Thus, it's the CRLB of $\hat{A}$, but multiplied by $\left(\frac{\partial g}{\partial \theta}\right)^2$.
- For example, in the above case of estimating $\alpha = A^2$, the CRLB becomes:

$$\operatorname{var}(\widehat{A^2}) \geqslant \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}.$$

CRLB - Transformation of parameters (cont.)

- However, the MVU *estimator* of A does not transform to the MVU estimator of A^2.
- This is easy to see by noting that it's not even unbiased:

$$E(\bar{x}^2) = E^2(\bar{x}) + \mathrm{var}(\bar{x}) = A^2 + \frac{\sigma^2}{N} \neq A^2$$

- The efficiency of an estimator is destroyed by a nonlinear transformation.
- The efficiency of an estimator *is however* maintained for affine transformations.

CRLB - Transformation of parameters (cont.)

- Additionally, the efficiency of an estimator is *approximately* preserved over nonlinear transformations *if the data record is large enough*.
- For example, $\hat{\alpha}$ above is *asymptotically* unbiased. Since $\bar{x} \sim \mathcal{N}(A, \sigma^2/N)$, it can be shown, that the variance $\mathrm{var}(\bar{x}^2) = \frac{4A^2\sigma^2}{N} + \frac{2\sigma^4}{N^2} \rightarrow \frac{4A^2\sigma^2}{N}$, as $N \rightarrow \infty$.
- Thus, the variance approaches the CRLB as $N \rightarrow \infty$, and $\hat{\alpha} = \hat{A}^2$ is *asymptotically efficient*.

CRLB - Vector Parameter

Theorem: CRLB - vector parameter
It is assumed that the pdf $p(\mathbf{x};\theta)$ *satisfies the "regularity" condition*

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0, \quad \forall \theta \tag{4}$$

where the expectation is taken with respect to $p(\mathbf{x};\theta)$. *Then, the covariance matrix of any unbiased estimator* $\hat{\theta}$ *satisfies:*

$$\mathbf{C}_{\hat{\theta}} - \mathbf{I}^{-1}(\theta) \geqslant 0 \tag{5}$$

CRLB - Vector Parameter (cont.)

where "$\geqslant 0$" is interpreted as positive semidefinite. The Fisher information matrix $\mathbf{I}(\theta)$ is given as

$$[\mathbf{I}(\theta)]_{ij} = -E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta_i \partial\theta_j}\right]$$

where the derivatives are evaluated at the true value of θ and the expectation is taken with respect to $p(\mathbf{x};\theta)$.
Furthermore, an unbiased estimator may be found that attains the bound in that $\mathbf{C}_{\hat{\theta}} = \mathbf{I}^{-1}(\theta)$ if and only if

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} = \mathbf{I}(\theta)(g(x) - \theta) \tag{6}$$

CRLB - Vector Parameter (cont.)

for some function g *and some* $m \times m$ *matrix* $\mathbf{I}$ $(m = \dim(\theta))$*. That estimator, which is the MVU estimator, is* $\hat{\theta} = g(\mathbf{x})$ *and the covariance matrix is* $\mathbf{I}^{-1}(\theta)$.

The special case of WGN

- Again, the important special case of WGN is easier than the general case.
- Suppose we have observed the data $\mathbf{x} = [x[0], x[1], \ldots, x[N-1]]$, and assume the model

$$x[n] = s[n] + w[n],$$

  where $s[n]$ depends on the parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_m]$ and $w[n]$ is WGN
- In this case the Fisher information matrix becomes

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial s[n]}{\partial \theta_i} \frac{\partial s[n]}{\partial \theta_j}$$

CRLB with vector parameters—Example

- Consider the sinusoidal parameter estimation seen earlier, but this time with A, f_0 and ϕ all unknown:

$$x[n] = A\cos(2\pi f_0 n + \phi) + w[n], \qquad n = 0, 1, \ldots, N-1,$$

with $w[n] \sim \mathcal{N}(0, \sigma^2)$. We can not derive an estimator yet, but let's see what are their bounds.

- Now the parameter vector θ is

$$\theta = \begin{pmatrix} A \\ f_0 \\ \phi \end{pmatrix}.$$

CRLB with vector parameters—Example (cont.)

- Because the noise is white Gaussian, we can apply the WGN form of CRLB from the previous slide (easier than the general form). The Fisher information matrix $\mathbf{I}(\theta)$ is a 3×3 matrix given by

$$[\mathbf{I}(\theta)]_{ij} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial s[n]}{\partial \theta_i} \frac{\partial s[n]}{\partial \theta_j}$$

- Let us simplify the notation by using $\alpha = 2\pi f_0 n + \phi$ when possible.

CRLB with vector parameters—Example (cont.)

$$
\begin{aligned}
[\mathbf{I}(\theta)]_{11} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial s[n]}{\partial A}\frac{\partial s[n]}{\partial A} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\cos^2\alpha \approx \frac{N}{2\sigma^2} \\
[\mathbf{I}(\theta)]_{12} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial s[n]}{\partial A}\frac{\partial s[n]}{\partial f_0} = -\frac{1}{\sigma^2}\sum_{n=0}^{N-1}A2\pi n\cos\alpha\sin\alpha = -\frac{\pi A}{\sigma^2}\sum_{n=0}^{N-1}n\sin 2\alpha \approx 0 \\
[\mathbf{I}(\theta)]_{13} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial s[n]}{\partial A}\frac{\partial s[n]}{\partial \phi} = -\frac{1}{\sigma^2}\sum_{n=0}^{N-1}A\cos\alpha\sin\alpha = -\frac{A}{2\sigma^2}\sum_{n=0}^{N-1}\sin 2\alpha \approx 0 \\
[\mathbf{I}(\theta)]_{22} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial s[n]}{\partial f_0}\frac{\partial s[n]}{\partial f_0} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(A2\pi n\sin\alpha)^2 \approx \frac{(2\pi A)^2}{2\sigma^2}\sum_{n=0}^{N-1}n^2 \\
&= \frac{(2\pi A)^2 N(N-1)(2N-1)}{12\sigma^2} \\
[\mathbf{I}(\theta)]_{23} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial s[n]}{\partial f_0}\frac{\partial s[n]}{\partial \phi} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}A^2 2\pi n\sin^2\alpha \approx \frac{\pi A^2}{\sigma^2}\sum_{n=0}^{N-1}n = \frac{\pi A^2 N(N-1)}{2\sigma^2} \\
[\mathbf{I}(\theta)]_{33} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial s[n]}{\partial \phi}\frac{\partial s[n]}{\partial \phi} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}A^2\sin^2\alpha \approx \frac{NA^2}{2\sigma^2}
\end{aligned}
$$

CRLB with vector parameters—Example (cont.)

- All approximations are based on the following triangular identities

$$
\begin{aligned}
\cos 2\alpha &= 2\cos^2\alpha - 1 \Rightarrow \cos^2\alpha = \frac{1}{2} + \frac{1}{2}\cos 2\alpha \\
\sin 2\alpha &= 2\sin\alpha\cos\alpha \Rightarrow \sin\alpha\cos\alpha = \frac{1}{2}\sin 2\alpha
\end{aligned}
$$

A sum of sines or cosines becomes negligible for large N.

CRLB with vector parameters—Example (cont.)

- Since the Fisher information matrix is symmetric, we have all we need to form it:

$$\mathbf{I}(\theta) = \frac{1}{\sigma^2}\begin{pmatrix} \frac{N}{2} & 0 & 0 \\ 0 & \frac{(\pi A)^2 N(N-1)(2N-1)}{3} & \frac{\pi A^2 N(N-1)}{2} \\ 0 & \frac{\pi A^2 N(N-1)}{2} & \frac{NA^2}{2} \end{pmatrix}$$

- Inversion with symbolic math toolbox gives the inverse:

$$\mathbf{I}^{-1}(\theta) = \sigma^2\begin{pmatrix} \frac{2}{N} & 0 & 0 \\ 0 & \frac{6}{A^2 N(N^2-1)\pi^2} & -\frac{6}{A^2(N+N^2)\pi} \\ 0 & -\frac{6}{A^2(N+N^2)\pi} & \frac{4(2N-1)}{A^2(N+N^2)} \end{pmatrix}$$

CRLB with vector parameters—Example (cont.)

- The CRLB theorem states that the covariance matrix $\mathbf{C}_{\hat{\theta}}$ of any unbiased estimator satisfies

$$\mathbf{C}_{\hat{\theta}} - \mathbf{I}^{-1}(\theta) \geqslant \mathbf{0},$$

  which means that the matrix on the left is positive semidefinite.

- In particular, all diagonal elements of a positive semidefinite matrix are non-negative, that is,

$$\left[\mathbf{C}_{\hat{\theta}} - \mathbf{I}^{-1}(\theta)\right]_{ii} \geqslant 0.$$

CRLB with vector parameters—Example (cont.)

- In other words, the diagonal elements of the inverse of Fisher information matrix are the bounds for the three estimators:

$$\mathrm{var}(\hat{\theta}_i) = [\mathbf{C}_{\hat{\theta}_i}]_{ii} \geqslant \left[\mathbf{I}^{-1}(\theta)\right]_{ii}$$

- The end result is, that

$$\begin{aligned}
\mathrm{var}(\hat{A}) &\geqslant \frac{2\sigma^2}{N} \\
\mathrm{var}(\hat{f}_0) &\geqslant \frac{6\sigma^2}{A^2 N(N^2-1)\pi^2} \\
\mathrm{var}(\hat{\phi}) &\geqslant \frac{4(2N-1)}{A^2(N+N^2)}
\end{aligned}$$

CRLB with vector parameters—Example (cont.)

- It is interesting to note, that the bounds are higher than when estimating the parameters individually.
- For example, the minimum variance of $\hat{f}_0$ is significantly higher than when the other parameters are known.
- The plot below shows the approximated lower bound in this case (straight line) compared to earlier bound for individual estimation of $\hat{f}_0$ (curve).

CRLB with vector parameters—Example (cont.)